European Journal of Neuroscience, Vol. 13, pp. 2004–2008, 2001

Purely temporal figure–ground segregation

Farid I. Kandil and Manfred Fahle
Institute of Human Neurobiology, University of Bremen, Argonnenstr. 3, 28211 Bremen, Germany

Keywords: age effects, Gestalt perception, human beings, psychophysics, segmentation

Abstract

Visual figure–ground segregation is achieved by exploiting differences in features such as luminance, colour, motion or presentation time between a figure and its surround. Here we determine the shortest delay times required for figure–ground segregation based on purely temporal features. Previous studies usually employed stimulus onset asynchronies between figure- and ground-containing possible artefacts based on apparent motion cues or on luminance differences. Our stimuli systematically avoid these artefacts by constantly showing 20 × 20 'colons' that flip by 90° around their midpoints at constant time intervals. Colons constituting the background flip in-phase whereas those constituting the target flip with a phase delay. We tested the impact of frequency modulation and phase reduction on target detection. Younger subjects performed well above chance even at temporal delays as short as 13 ms, whilst older subjects required up to three times longer delays in some conditions. Figure–ground segregation can rely on purely temporal delays down to around 10 ms even in the absence of luminance and motion artefacts, indicating a temporal precision of cortical information processing almost an order of magnitude lower than the one required for some models of feature binding in the visual cortex [e.g. Singer, W. (1999), *Curr. Opin. Neurobiol.*, **9**, 189–194]. Hence, in our experiment, observers are unable to use temporal stimulus features with the precision required for these models.

Introduction

Analysis of a complex visual image requires the solution of two related problems, segregating figure from ground and joining elements into a figure (grouping). Gestalt psychologists found that grouping of elements is based on similarity in features (such as luminance and colour), spatial neighbourhood or 'common fate', whilst dissimilarity in these features, distance, and different fates lead to segregation (Köhler, 1947). As a special case of the law of common fate, simultaneous changes lead to grouping whereas asynchronous changes lead to segregation. The minimal temporal delay between changes of figure and ground required for segregation hence demarcates the range of perceived simultaneity (Pöppel, 1997). It is a basic constant of the visual system and allows estimation of the temporal precision of human cortical information processing. Two types of perceptual tasks served to measure time-based segregation.

In the first, an array is filled with a homogeneous group of flickering dots, lines or squares (Fahle, 1993; Kiper, Gegenfurtner & Movshon, 1996; Leonards, Singer & Fahle, 1996; Leonards & Singer, 1998; Rogers-Ramachandran & Ramachandran, 1998; Usher & Donnelly, 1998; Forte, Hogben & Ross, 1999). The only difference between figure and ground is that elements of the figure appear and disappear with a time (phase) delay. Even at flicker frequencies of $\geqslant$ 30 Hz and delays around 10–20 ms, subjects were usually able to detect the targets. However, these displays contain frames showing only the figure whilst others present the background alone. It has been argued that a subject able to isolate a single frame performs the figure–ground segregation based on luminance differences: the background will be dark in a frame containing only the figure (Lee & Blake, 1999a). Another possible artefact is apparent motion: one frame displays the dots constituting the figure whilst the successive one displays those of the ground with a certain delay, possibly inducing apparent motion across the border between figure and ground, and thus demarcating this border. Therefore these stimuli cannot exclude a detection of the target by first-order motion or fast first-order luminance detectors.

The second paradigm (Lee & Blake, 1999a) avoids artefacts based on first-order luminance and first-order motion cues by showing a set of 'windmills' rotating in either direction. At random points in time, the windmills in the target area, or else those in the surround, change direction. However, as Adelson & Farid (1999) criticised, long runs in the same direction result in lower local stimulus contrast whereas a stop-and-return results in a transient high contrast. [Lee & Blake (1999b) subsequently supplied additional evidence against the suspicion that these long runs or stop-and-go sequences might have been the basis of shape detection in their experiment.] Lee and Blake's temporal protocol, unlike our experiments, does not allow precise statements about the shortest temporal differences sufficient to perform the segregation.

In this paper, we present new displays constructed to prevent figure–ground segregation based on luminance differences and first-order motion that allow investigation of minimum delays required to discriminate figure from ground by means of purely temporal cues.

Materials and methods

Stimuli

Stimulus displays used in expts 1 and 2 are shown in Fig. 1A and B. Each of the four frames presents 20 × 20 colons which flip around their theoretical midpoints by 90° after every second display. Colons in the target area flip between displays 1 and 2 and between displays 3

Correspondence: Farid Kandil, as above.
E-mail: kandil@uni-bremen.de

Received 30 January 2001, revised 21 March 2001, accepted 26 March 2001

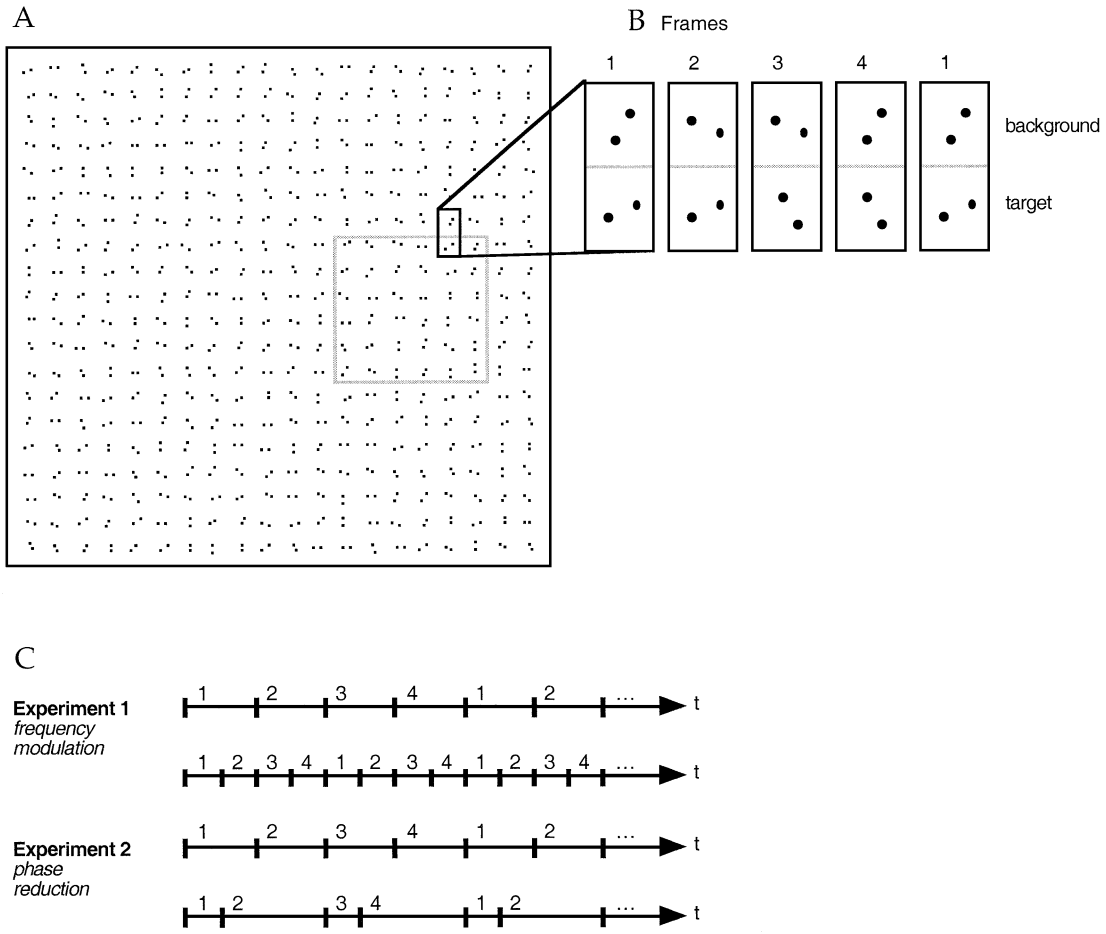

FIG. 1. (A) Displays are filled with 20 × 20 colons which flip by 90 ° around their (imaginary) midpoints after every other frame. (B and C) The colons forming the target and those forming the background flip at different points in time: in expt 1, flip frequency increases between blocks from 4.2 to 33.3 Hz, while target and ground continue flipping exactly in counter-phase. In expt 2, flip frequency is constant at a low level of 4 Hz while phase angle between target and ground is reduced.

and 4, whilst surround elements flip between displays 2 and 3 and between displays 4 and 1. The variable in expts 1 and 2 is the amount of time each of the four possible frames is continuously presented. Displays were calculated by an Apple Macintosh PowerPC and presented on an analogue monitor (HP 1332, P11 phosphor) via fast 16-bit D/A converters. One complete frame could be displayed within 5 ms, i.e. the maximum possible frame rate was 200 Hz. At lower frame rates, each frame was intensified more than once. Stimulus dots had a luminance of ≈ 140 cd/m^2 on a background of around 1.2 cd/m^2 (contrast 98%). The room was lit at approximately 6 lux. At a viewing distance of 30 cm the displays were 14.6° wide and high, virtual grid width was 45 arcmin, a colon measured 12 arcmin from dot to dot and a single dot had a diameter of 3.4 arcmin. Virtual midpoints of the colons were jittered around the regular grid position, both horizontally and vertically by either –3, 0 or +3 arcmin, and colons' starting orientations differed between 0° and 150° from the horizontal, in 30 ° steps. Stimuli were shown for a maximum of 3 s in all experiments.

In expts 1 and 2, the target area was 6 × 6 colons in size, located at one of four fixed locations, at the middle of either the left, right, top or bottom half, starting on the second row/column from the border. In a four-alternative forced choice (4-AFC) paradigm, subjects had to localize the targets and to report their position via a four-stroke keyboard.

Data analysis

For each subject and each temporal condition we calculated the percentage of correct answers across the number of trials tested, 48 in expt 1 and 24 in expts 2 and 4. In pilot studies, we did not find any single frequency suited to compare all subjects without ceiling and floor effects. Therefore, the threshold delay $\Delta t_{62.5\%}$ was determined or linearly interpolated individually for each subject. Thresholds indicate the minimum delay required to reach a level of 62.5% correct responses (midway between perfect and chance level). Comparison between the age groups was conducted using the Fisher–Pitman test.

Subjects

Twenty-four subjects in four different age classes, without any psychiatric, neurological or ophthalmologic history, participated in expts 1, 2 and 4. Another group of five subjects participated in expt 3. All had normal or corrected-to-normal visual acuity. Approval for this type of experiments had been obtained by both the Tübingen and the Bremen review and approval committees.

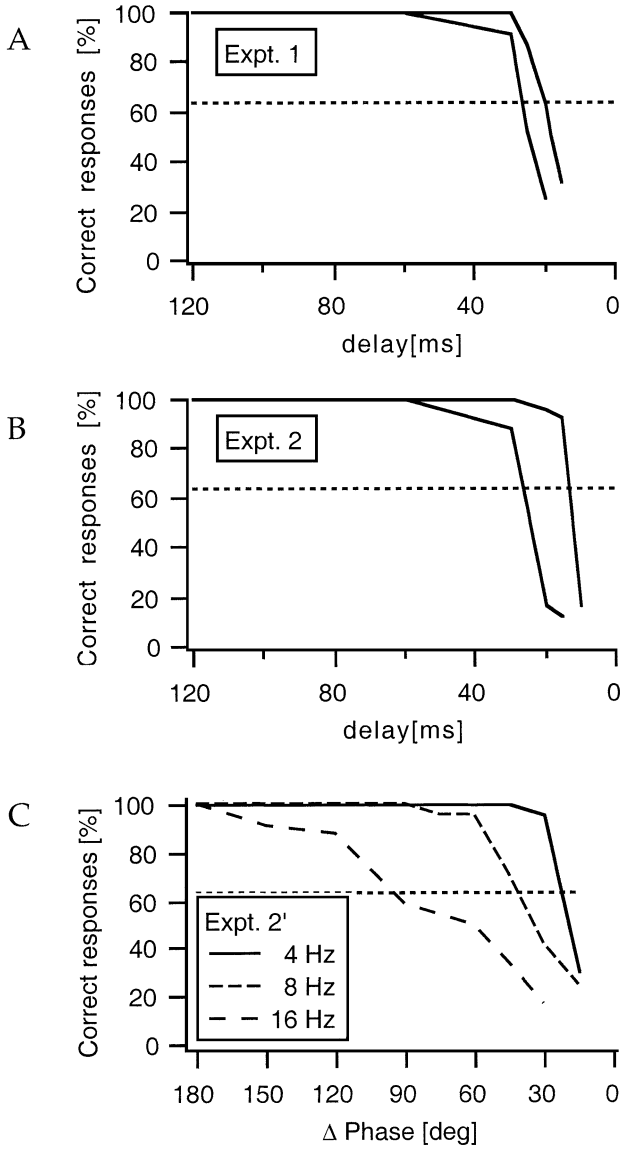

FIG. 2. Results for two subjects in (A) expt 1 and (B) expt 2. Abscissa in both diagrams indicates delay in ms. Subjects start at a frequency of 4 Hz and with figure and ground flipping at counter-phase to each other. As the frequency increases (A) or phase delay decreases (B) (reducing delays), performance sinks from a perfect level to chance (25%). Thresholds are defined as the 62.5%-quantile (dotted lines). (C) Results for one additional subject from the pilot study for expt 2 where different cycle frequencies (4, 8 and 16 Hz) were used. The graphs demonstrate that performance decreased monotonically with the reduction of phase angle Δ.

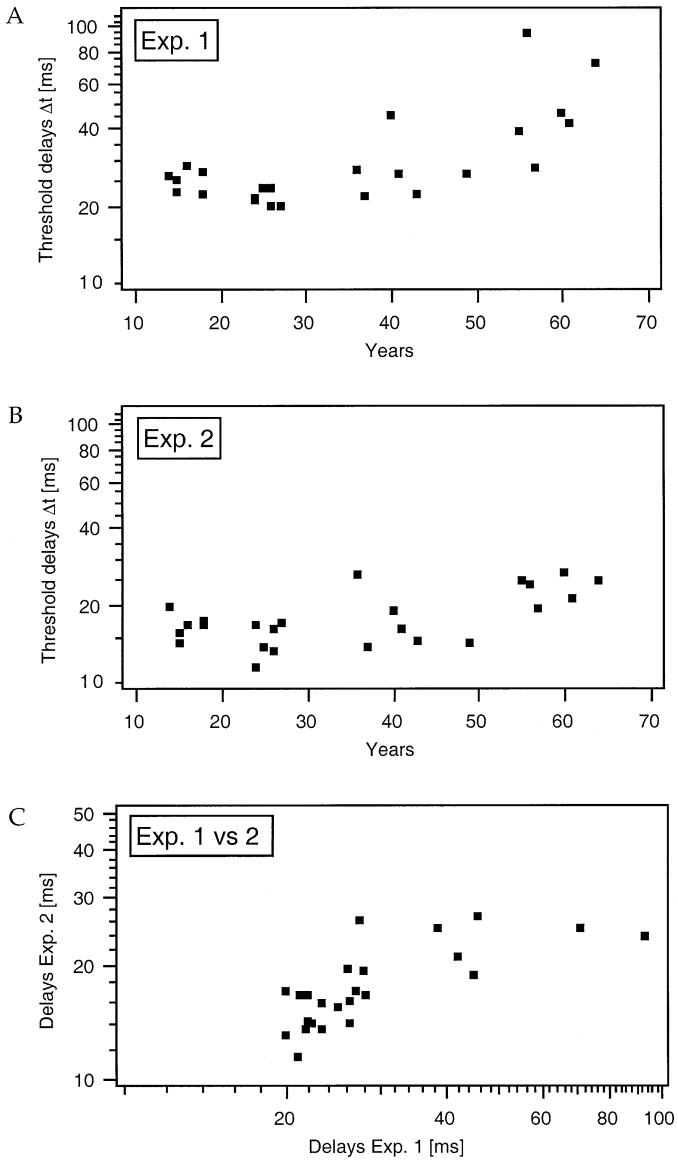

FIG. 3. (A and B) Individual threshold delays (Δt, in ms) obtained from expts 1 (frequency modulation) and 2 (phase reduction) as functions of age, and (C) their correlation. Thresholds in expt 1 are rather homogeneous for younger subjects (< 40 years) but increase thereafter. (B and C) Thresholds in expt 2 are lower by a factor of 2–3 but yield a comparable age-dependency.

Results

Experiment 1

Cycle frequency was increased from 4.2 to 33.3 flips per second (Hz) by reducing the presentation times of all frames uniformly from 120 to 5 ms by reducing the number of intensifications. Here, target and background flipped in counter phase to each other (cf the upper half of Fig. 1C).

Complete data sets for two subjects (Fig. 2A) and individual threshold delays (Δt) for all subjects (Fig. 3A) show that subjects between 20 and 35 years of age (group II) yielded best results (i.e. the highest maximum stimulus frequencies and thus lowest minimum Δt). They were able to detect the target even at frequencies of 22.7 Hz (period length 44 ms, Δt 22 ms). Younger (group I, 15–20 years) and older (group III, 35–50 years) subjects performed slightly worse (18–20 Hz, Δt 25–28 ms), whereas most subjects older than 50 years (group IV) needed longer delays (9.43 Hz, Δt 53 ms). The (alpha-adjusted) Fisher–Pitman test revealed significant differences between the younger groups (I, II, III) and the older group (IV) (Table 1).

Most subjects reported that they had been able to follow the temporal modulations at 4 and 8 Hz and only one reported the ability to detect them even at 16 Hz. The two subjects with the highest threshold delays experienced problems following even modulations of 8 Hz.

TABLE 1. Figure–ground discrimination and on-time duration at the CFF

Group	Temporal figure–ground discrimination (ms)				On-time durations at the CFF
	Experiment 1	*P*	Experiment 2	*P*	
I	25.14 ± 1.04	0.0139	16.53 ± 0.72	0.0025	17.01 ± 0.55
II	21.37 ± 0.56	0.0102	14.59 ± 0.92	0.0017	18.86 ± 1.10
III	28.16 ± 3.51	0.0244	17.13 ± 1.95	0.0143	22.11 ± 1.00
IV	52.95 ± 9.90		23.31 ± 1.12		21.13 ± 1.10

Data are presented as means ± SEM. CFF, critical flicker fusion frequency. *P*-values from the Fisher–Pitman are shown for age groups I–III against group IV. On-time durations at the critical flicker fusion frequency (CFF) in ms.

Experiment 2

Some authors used phase reduction instead of frequency increase and obtained clearly lower temporal thresholds than we did in expt 1 (Fahle, 1993; Leonards, Singer & Fahle, 1996; Leonards & Singer, 1998), and Usher & Donnelly (1998) found best performance when they presented both target and background only once, as opposed to several times. Therefore, we repeated expt 1 with the paradigm of phase reduction (lower half of Fig. 1C).

At a low frequency of 4.2 flips per second (which again corresponds to a total presentation time of 240 ms for each pair of frames), presentation times of frames 1 and 3 were reduced from 120 to 5 ms, and hence presentation times of frames 2 and 4 were prolonged from 120 to 235 ms.

Figure 2B shows psychometric functions for two subjects and Fig. 3B displays individual threshold delays Δt for all subjects. Best results were ≈11–13 ms, poorest at ≈25–27 ms and hence shorter than in expt 1 by a factor of 1.5–4 (Fig. 3C).

Experiment 3

Experiment 3 tested whether first-order apparent motion between neighbouring colons belonging to figure vs. ground was strong enough to subserve target detection based on the time differences between the flips in figure vs. ground. First-order apparent motion is based on luminance displacements. Our paradigm is based on strictly local displacements of luminance in every flip, and hence there is potential for apparent motion in every flip. However, we argue that the motion is strictly local with distances between neighbouring colons being ≈50 arcmin, four times longer than distances between flip positions, and that subjectively no motion is experienced between colons.

To test more rigorously the influence of apparent motion across borders between figure and ground we presented displays consisting of targets and grounds in the form of stripes of variable width. Smaller stripe width leads to longer overall borders between the stripes than does larger stripe width, as is most evident for the case of a stripe width corresponding to half of the stimulus width. In this case, there is only one single border across which apparent motion between neighbouring columns might occur. The rationale is that if this hypothetical apparent motion across the borders between figure and ground was the major cue for segmentation, detection of these borders and hence of the stripes' orientation should improve with the total length of borders: hence with smaller stripe width there is a better signal-to-noise ratio because of the longer borders.

Displays presented 12 × 12 colons which were divided into either 12, 6, 4 or 2 rows or columns with widths of 1, 2, 3 or 6 elements, respectively, so that always half of the colons defined the 'target' (set 1) and the other half the 'background' (set 2). Five subjects had to decide whether the orientation of the resulting grating was horizontal or vertical; this was a two-alternative forced-choice (2-AFC). Each subject participated in 160 trials at a stimulus frequency at which ≈75% of the trials could be answered correctly to ensure maximum sensitivity.

The four conditions in five subjects tested at each individual's critical frequency did not differ significantly. (Randomization test for dependent samples: $P > 0.10$).

Experiment 4

Limits of temporal resolution for figure–ground segregation, especially in expt 1, could be a consequence of critical flicker fusion frequency. If presentations of subsequent stimuli followed each other faster than the critical flicker frequency (CFF), observers might no longer be able to detect phase differences between the stimuli: it would not be too surprising if observers were unable to detect phase differences between stimuli whose (flicker) frequencies they cannot resolve. To test this possible artefact, we performed a modified test of CFF. Here, displays differed strongly from the ones in the previous experiments in that only four single dots were presented, in the form of a cross. Subjects had to decide which of the four dots was flickering (4-AFC). Frequency increased stepwise from 4.2 Hz (on-phase of the flickering dot, 120 ms) to 50 Hz (on-phase 10 ms). As before, individual thresholds were defined as the 62.5% quantiles. Luminance and dot size were kept on the same level as in the other experiments to ensure comparability.

Critical flicker fusion frequencies ranged between 22 and 32 Hz between subjects, corresponding to durations of the on-phases between 23 and 16 ms (Table 1).

Discussion

The results of expts 1 and 2 confirm that subjects were able to perform perceptual grouping on the basis of strictly temporal features. In paradigms such as expt 1, grouping requires delays (Δt) between target and ground of ≈20 ms. In contrast, at constant flip frequency and variable phase angle, minimum delays of ≈11–13 ms are sufficient, agreeing well with the finding that thresholds are lowest for single rather than repetitive presentations of some stimuli (Usher & Donnelly, 1998).

Experiment 3 demonstrated that the number of the rows and columns (and hence the total length of the border) did not play a significant role, another indication that segregation was not based on apparent motion occurring at the borders between target and ground (sets 1 and 2).

On-phases of the critical flicker frequencies obtained in expt 4 varied between 16 and 23 ms, in the same range as the minimum delays in expt 2 although 1.5× lower than those of expt 1. This means that the delay between presentations of a single dot necessary to perceive it as flickering at a fixed position was sufficient to discriminate between figure and ground when dots were spatially separated.

Comparing the results of expts 1 (constant phase) and 2 (variable phase), two possible reasons for the difference between their results spring to mind: the first is that observers may be able to use the prolonged complementary delays such as between frames 2 and 3 and frames 4 and 1 in the lowest row of Fig. 1C. If this were true the psychometric functions of observers should be nonmonotonic, with at least slightly better performances for slightly out of counter-phase presentations, because the presentations contain longer complementary delays than the counter-phase presentations. We did not find such

psychometric functions in the pilot studies for expt 2 (Fig. 2C). Hence the important parameter seems to be the minimal delay: as mentioned above, once two frames are fused, the subsequent delay cannot easily be used to detect which of the two frames was presented later than its partner. Indeed, the fourth experiment clearly showed CFF to roughly correspond to the best results obtained in the first experiment. The second possible explanation for the better results in the phase-variable condition (expt 2) is the fact that there, the presentation frequency (4.2 Hz) was much lower than the CFF and hence stimulus onsets were clearly detectable although they blurred close to the CFF. We favour this second interpretation.

Figure–ground segregation in the experiments of Lee and Blake (1999a) occurred even at high mean reversal frequencies. Our results indicate that, under these conditions, observers' decisions were based on the lower stimulus frequencies always present in their displays.

Observers aged between 15 and 40 years detected the target best (i.e. with the shortest delays), whereas subjects above ≈ 50 years needed two to three (in two cases even more) times longer delays in expt 1. This age-related deterioration is not highly correlated with any of the visual functions tested (CFF: $r = 0.368$; visual acuity: $r = -0.242$) apart from the form-from-motion detection task (DDVT; Wist *et al*., 2000: $r = -0.764$). Both tests, the DDVT as well as our test, require motion detection and discrimination. Thus, the high correlation is not surprising. The age effect was much smaller under the phase-variable condition. These results indicate that the temporal resolution as measured with our test suffers more from ageing than does the spatial resolution.

Experiments 1 and 2 demonstrate that subjects are able to segment figure from ground solely on the basis of temporal delays. The target cannot be detected in a single display and does not contain luminance differences between frames, thus excluding the usage of first-order motion detectors for figure–ground discrimination. Experiment 3 ensured that hypothetical first-order apparent motion signals emerging at the target borders are not strong enough to allow identification of the target. Luminance, motion of the stimulus dots and their flicker frequency are identical within both the target and the surround. The only difference between figure and ground is the point in time when motion takes place. Thus figure–ground segmentation in our experiments cannot be subserved either by first-order motion detectors or by first-order flicker detectors, but relies on purely temporal or more complex, that is higher-order, spatio-temporal detectors, analysing the output of (elementary) motion or flicker detectors. These detectors yield thresholds around one hundredth of a second simultaneously at many positions of the visual field. These short delays, astonishing as they may appear, are still almost an order of magnitude higher than the ones required by some models of object formation based on feature binding via synchronization of cortical action potentials (Singer, 1999). A possible explanation is that our stimuli may be unable to synchronize the cortical neurons with sufficient precision, i.e. to drive the internal code. This question can only be decided on the basis of electrophysiological experiments.

Acknowledgements

Supported by Deutsche Forschungsgemeinschaft (SFB 517).

Abbreviations

2-AFC, two-alternative forced-choice; 4-AFC, four-alternative forced-choice; CFF, critical flicker frequency; Δt, threshold delay (ms); Hz, flips per second.

References

Adelson, E.H. & Farid, H. (1999) Filtering reveals form in temporal structured displays. *Science*, **286**, 2231a.

Fahle, M. (1993) Figure-ground discrimination from temporal information. *Proc. R. Soc. Lond. B*, **254**, 199–203.

Forte, J., Hogben, J.H. & Ross, J. (1999) Spatial limitations of temporal segmentation. *Vision Res.*, **39**, 4052–4061.

Kiper, D.C., Gegenfurtner, K.R. & Movshon, A. (1996) Cortical oscillatory responses do not affect visual segmentation. *Vision Res.*, **36**, 539–544.

Köhler, W. (1947). *Gestalt Psychology*. Meridian Press, New York.

Lee, S.H. & Blake, R. (1999a) Visual form created solely from temporal structure. *Science*, **284**, 1165–1168.

Lee, S.H. & Blake, R. (1999b) Response to: Filtering reveals form in temporal structured displays. *Science*, **286**, 2231a–2232a.

Leonards, U. & Singer, W. (1998) Two segmentation mechanisms with differential sensitivity for color and luminance contrasts. *Vision Res.*, **38**, 101–109.

Leonards, U., Singer, W. & Fahle, M. (1996) The influence of temporal phase differences on texture segmentation. *Vision Res.*, **36**, 2689–2697.

Pöppel, E. (1997) A hierachical model of temporal perception. *Trends Cognit. Sci.*, **1**, 56–61.

Rogers-Ramachandran, D.C. & Ramachandran, V.S. (1998) Psychophysical evidence for boundary and surface systems in human vision. *Vision Res.*, **38**, 71–77.

Singer, W. (1999) Time as coding space? *Curr. Opin. Neurobiol.*, **9**, 189–194.

Usher, M. & Donnelly, N. (1998) Visual synchrony affects binding and segmentation in perception. *Nature*, **394**, 179–182.

Wist, E.R., Schrauf, M. & Ehrenstein, W.H. (2000) Dynamic vision based on motion contrast: changes with age in adults. *Exp. Brain. Res.*, **134**, 295–300.